~~SECRET~~

23 March 1970

MEMORANDUM FOR: Mr. John K. Vance

SUBJECT: Cost-Effectiveness Analysis of AEGIS Indexing Depth

You have raised the question as to whether the present AEGIS depth of indexing is "optimum" from a cost-effectiveness viewpoint and whether it is possible to undertake some study of the cost-effectiveness of AEGIS indexing. I suggested a possible methodology in our last meeting and will attempt to formalize this below.

First, we must recognize what we can expect to happen to the effectiveness of AEGIS if we increase the average number of items assigned. We would expect that the average recall of the system would increase and that the average precision would decrease. Since my evaluation tended to indicate that AEGIS was somewhat low on recall but high on precision (which can usually be improved by the buffering operation) on the surface it seems desirable to increase indexing depth. But we really need to estimate how many additional terms we need to assign, on the average, to raise recall X%. At the same time we need to estimate what effect this will have on indexing costs. Then we can balance expected increase in costs against expected increase in effectiveness, allowing a management decision to be made on whether or not the anticipated increase in effectiveness justifies the increased costs.

In my previous evaluation of AEGIS, we have some preliminary data on the effect of present indexing exhaustivity (depth) on retrieval performance. "Lack of exhaustivity" contributed to 2/18 failures (11%) in the real-life searches and 5/28 failures (18%) in the synthetic searches.

~~SECRET~~

Let us say, then, that 10-20 % of all recall failures are due to shallow indexing. If we increased the exhaustivity of indexing X % we would expect to be able to eliminate this group of failures and to raise the over-all recall average from about 50 % to 55 % or 60 %. On the surface this seems like a fairly insignificant improvement - at least it would be hard to justify on cost-effectiveness grounds if it meant a substantial increase in indexing costs. However, increased exhaustivity of indexing might actually raise recall more than an average of 5-10 %. This would be caused by the fact that the increase in exhaustivity would compensate for some of the other failures by providing a kind of "fail-safe" mechanism. For example, I may have attributed most of the recall failures in a search to the fact that the searcher did not use term A, which appeared highly appropriate to the request (i.e., it was classed as a searching failure). With more exhaustive indexing some of the additional terms assigned might have matched the terms that the searcher did use. Thus, increased exhaustivity (depth) might in fact lead to a 15-25 % improvement in average AEGIS recall performance, rather than a 5-10 % increase which appears to be the potential on the surface.

The procedures for conducting the test specifically on cost-effectiveness of exhaustivity would be as follows:

1. Take the 22 finished intelligence documents used in the earlier test and gather once more the documentation used by the authors of these reports. This time, however, we would identify all documents cited that should be in the AEGIS data base rather than just a sample.

2. Go back to the search printout for each of the 22 searches conducted (I have these) and determine how many of the cited documents, in the AEGIS data base, were retrieved (i.e., establish an expanded recall figure for each search).

3. Take each nonretrieved document and have it reindexed at varying levels of exhaustivity under controlled conditions (including timing). Three levels could be tried:

(1) shallow (present level)
(2) intermediate } we will need to specify what
(3) full } we mean by each level

4. Take this new indexing and compare it manually with the strategies used in the original searches.

5. Measure the improvement in recall allowed by the increased exhaustivity and calculate indexing costs associated with each level.

6. Balance costs against expected recall improvement to allow a management decision to be made.

NOTES

1. This is a skeleton of a procedure only. The methodology obviously requires refining.

2. I am not necessarily advocating an increase in AEGIS depth, but merely pointing out that the probable effects can be estimated.

3. This study does not require contacting any analysts outside of CRS. The necessary contact has already been made for purposes of the previous study.

4. I could begin work on such a study, and possibly complete it, within the scope of my present contract - if you feel it is worthwhile.

5. We can discuss the methodology in greater detail at your leasure.

25X1A9a

UNCLASSIFIED | CONFIDENTIAL | SECRET

CENTRAL INTELLIGENCE AGENCY

OFFICIAL ROUTING SLIP

TO	NAME AND ADDRESS	DATE	INITIALS
1	~~ADDI~~		
2			
3			
4			
5			
6			

	ACTION		DIRECT REPLY		PREPARE REPLY
	APPROVAL		DISPATCH		RECOMMENDATION
	COMMENT		FILE		RETURN
	CONCURRENCE		INFORMATION		SIGNATURE

Remarks: Ed: Herew ith various think pieces from [redacted] 25X1A9a which indicate that an interesting and useful dialog is underway concerning many facets of the AEGIS indexing system.

I have discussed the current [redacted] 25X1A9a project with you previously. It has been coordinated with [redacted] 25X1A9a We expect some results from it in the near future and will talk in more detail to your questions (attached).

The Guide to Aegis, also done by [redacted] 25X1A9a has been reviewed by OCI and OER. It seems to me to be an excellent piece of work and we are ready now to plan its publication in an attractive format.

John

FOLD HERE TO RETURN TO SENDER

FROM: NAME, ADDRESS AND PHONE NO.	DATE
D/CRS	24 Mar 70

UNCLASSIFIED | CONFIDENTIAL | SECRET

(40)

FORM NO. 2-61 237 Use previous editions

U.S. GOVERNMENT PRINTING OFFICE : 1961 O—587282